7.9 - 1 0.2 5 1.

NASA CR-
160285

LACIE-00633
JSC-14849

NASA NOAA USDA

LACIE TRANSITION PROJECT

(E79-10251) LARGE AREA CROP INVENTORY EXPERIMENT (LACIE). RESEARCH PLAN FOR DEVELOPING AND EVALUATING CLASSIFIERS (Lockheed Electronics Co.) 20 p HC A02/MF A01 CSCL 02C G3/43

N79-31711

Unclas
00251

RESEARCH PLAN FOR DEVELOPING AND EVALUATING CLASSIFIERS

NASA

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas 77058

JULY 1979

1. Report No. JSC-14849	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Research Plan for Developing and Evaluating Classifiers		5. Report Date April 1979
		6. Performing Organization Code LEC-13300
7. Author(s) C. B. Chittineni, Lockheed Electronics Company, Inc.		8. Performing Organization Report No.
		10. Work Unit No. 63-1827-3705
9. Performing Organization Name and Address Lockheed Electronics Company, Inc. Systems and Services Division 1830 NASA Road 1 Houston, Texas 77058		11. Contract or Grant No. NAS 9-15800
		13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Technical Monitor: J. D. Erickson/SF3		14. Sponsoring Agency Code

15. Supplementary Notes

LACIE-00633

16. Abstract

This plan describes proposals for developing and evaluating classifiers for analysis of data acquired by Landsat. Two classifiers were selected from a parametric family, and one classifier was selected from a nonparametric family. A leave-one-out method is proposed for effectively using type 1 dots (i.e., getting rid of type 2 dots) for estimating the biases, and computational expressions are presented for implementing it with the selected classifiers. The incorporation of contextual information with the classifiers is investigated, and contextual algorithms with a single parameter are described. A variance reduction factor is selected as a measure of performance of a machine-processing algorithm. It is proposed to compute the factor for all selected segments and for the selected classifiers with and without context and to compare the variance reduction factors with the ones obtained from the current Procedure 1 and with the ones obtained by using different clustering algorithms in proportion estimation. It is also proposed to investigate methods of handling imperfections in the labels.

17. Key Words (Suggested by Author(s))		18. Distribution Statement	
classifier evaluation, computational expressions, label imperfections, leave-one-out method, proportion estimation, research plan, variance reduction factor			
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 15	22. Price*

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75) NASA — JSC

ii

JSC-14849

RESEARCH PLAN FOR DEVELOPING AND EVALUATING CLASSIFIERS

APPROVED BY

R. P. Heydorn

R. P. Heydorn
Research, Test, and Evaluation Branch

Jon D. Erickson

J. D. Erickson, Chief
Research, Test, and Evaluation Branch

Earth Observations Division

Space and Life Sciences Directorate

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

April 1979 LEC-13300

JSC-14849

RESEARCH PLAN FOR DEVELOPING AND EVALUATING CLASSIFIERS

Job Order 73-705

BY

C. B. Chittineni

Prepared By

Lockheed Electronics Company
Systems and Services Division
Houston, Texas

Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

April 1979

LEC-13300

iv

CONTENTS

1. INTRODUCTION

A criterion used for assessing the performance of machine-processing algorithms during the Large Area Crop Inventory Experiment (LACIE) was the variance reduction factor. With the current processing procedure, Procedure 1, the observed variance reduction factor is high — around 0.75. This high value may be attributed to the low recognition accuracy of the current classification methods and to the need for estimating a large number of parameters. In addition, current procedures do not use scene spatial information in the classifiers; they use a separate set of labeled patterns (type 2 dots) for bias correction.

This plan proposes to evaluate classifiers that have significantly fewer parameters to estimate and use type 1 dots more effectively for estimating the biases. In particular, it is proposed to evaluate linear and piecewise linear classifiers and use leave-one-out methods for estimating the biases directly from type 1 dots.

2. SAMPLE SIZE AND DIMENSIONALITY

The error rates and hence the variance reduction factor estimates are a function of a particular sample used in the estimation. The expected values of the errors for a linear classifier can be derived as a function of training sample size and dimensionality (refs. 1 and 2). It has been observed that the minimum required ratio of number of samples to dimensionality per class is 3. Hence, this ratio will be used in these evaluation experiments. For 16-dimensional data then, the required number of samples per class is 48.

3. EVALUATION PROCEDURE

It is proposed to use 25 segments in the evaluation. The rationale for this number is given in reference 3, and the details of the particular segments selected are given in table 1. The variance reduction factor R will be used as a criterion for evaluation. The factor R is defined as

TABLE 1.— SEGMENTS TO BE USED IN THE EVALUATION OF CLASSIFIERS

Segment	Location (county, state)	Type of wheat	Proportion of small grains
1005	Cheyenne, Colo.	Winter	0.347
1032	Wichita, Kanas.	Winter	.386
1033	Clark, Kans.	Winter	.095
1853	Ness, Kans.	Winter	.303
1861	Kearny, Kans.	Winter	.353
1512	Clay, Minn.	Spring	.337
1520	Big Stone, Minn.	Spring	.308
1544	Sheridan, Mont.	Spring	.383
1739	Teton, Mont.	Mixed	.244
1582	Hays, Neb.	Winter	.194
1604	Renville, N. Dak.	Spring	.524
1606	Ward, N. Dak.	Spring	.329
1648	Bowman, N. Dak.	Spring	.379
1661	McIntosh, N. Dak.	Spring	.410
1902	McKenzie, N. Dak.	Spring	.086
1231	Jackson, Okla.	Winter	.741
1242	Canadian, Okla.	Winter	.472
1367	Major, Okla.	Winter	.540
1677	Spink, S. Dak.	Spring	.341
1690	Kingsbury, S. Dak.	Spring	.213
1803	Shannon, S. Dak.	Winter	.011
1805	Gregory, S. Dak.	Mixed	.158
1056	Moore, Tex.	Winter	.226
1059	Ochiltree, Tex.	Winter	.445
1060	Sherman, Tex.	Winter	.231

$$R = \frac{\sum_{i=1}^{M} \left(\frac{N_i}{N}\right) P_i (1 - P_i)}{P(1 - P)}$$

where

M is the total number of classes,

N is the total number of picture elements (pixels),

N_i is the total number of pixels in class i,

P is the overall proportion of small grains, and

P_i is the proportion of small grains in class i.

The proportion estimation procedures proposed for implementation with and without context are as follows. The estimated probability of occurrence of wheat is

$$\hat{P}(W) = \sum_{i=1}^{M} P(W|i)P(i)$$

where

$\hat{P}(W)$ is the estimated probability of occurrence of wheat;

$P(i)$ is the classifier estimated probability of occurrence of class i; and

$P(W|i)$, given the classifier decision as class i, is the probability of occurrence of wheat.

The $P(i)$'s are estimated in the usual way. For estimation $P(W|i)$, the leave-one-out method is used for the selected classifiers with and without context. The $P(W|i)$ is estimated from type 1 dots. Suppose that there are N_{SG} patterns of small grains and N_0 patterns of other in type 1 dots, where $N = N_{SG} + N_0$, the total number of type 1 dots. Using the leave-one-out method, suppose the following situation:

$N_{SG,SG}$ -- $N_{SG,SG}$ of N_{SG} labeled by the analyst as small grains and counted as small grains by the leave-one-out method;

$N_{SG,0}$ — $N_{SG,0}$ of N_{SG} labeled by the analyst as small grains and counted as other by the leave-one-out method;

$N_{0,SG}$ — $N_{0,SG}$ of N_0 labeled by the analyst as other and counted as small grains by the leave-one-out method; and

$N_{0,0}$ — $N_{0,0}$ of N_0 labeled by the analyst as other and counted as other by the leave-one-out method;

where

$$N_{SG} = N_{SG,SG} + N_{SG,0}, \; N_0 = N_{0,SG} + N_{0,0}$$

Then $P(W|i)$ is estimated as follows:

$$P(W|SG) = \frac{N_{SG,SG}}{N_{SG,SG} + N_{0,SG}}$$

and

$$P(W|0) = \frac{N_{SG,0}}{N_{SG,0} + N_{0,0}}$$

The variance reduction factor R will be computed for all the segments and for the selected classifiers with and without context. Based on R, the classification procedures will be compared with the current Procedure 1 and with the other cluster-based Procedure 1 being tested (ref. 3). Biases calculated from type 1 dots will be compared with those obtained from type 2 dots for all the segments.

4. USE OF THE LEAVE-ONE-OUT METHOD FOR ESTIMATING THE BIASES DIRECTLY FROM TYPE 1 DOTS

This section justifies the use of the leave-one-out method for estimating the biases directly from type 1 dots. Let Θ_1 be a set of parameters of the distributions used to design the classifier, and let Θ_2 be a set of parameters of the distributions of the patterns used to test the performance of the classifier.

Let $\varepsilon(\Theta_1, \Theta_2)$ be the resulting error when a classifier is designed on a set of patterns from distributions with parameters Θ_1 and tested on a set of patterns from distributions with parameters Θ_2. Let Θ and $\hat{\Theta}$ be the set of true parameters and its estimate. The $\hat{\Theta}$ is a random vector and depends on particular sample used in estimating it. Let $\hat{\Theta}_N$ be a particular value of $\hat{\Theta}$. Then,

$$\varepsilon(\Theta, \Theta) \leq \varepsilon(\hat{\Theta}_N, \Theta)$$

Taking expectations on both sides, one obtains

$$\varepsilon(\Theta, \Theta) \leq E[\varepsilon(\hat{\Theta}_N, \Theta)]$$

One of the ways of estimating the quantity on the RHS is using the leave-one-out method. This method is as follows. If there is a total of N patterns, leave out one pattern, design the classifier on remaining (N - 1) patterns, and test on the pattern that is left out. Repeat this procedure N times, each time leaving a different pattern. The estimated errors will be the estimates of the omission and commission errors and will be used in this evaluation procedure for bias correction.

5. LINEAR CLASSIFIERS

The linear classifiers selected for evaluation are the (a) Fisher classifier, which is parametric (refs. 4 and 5); (b) error correction classifier, which is nonparametric (refs. 6 and 7); and (c) classifier with uncertain labels, which is parametric. This section briefly describes these classifiers and presents computationally efficient methods for the use of the leave-one-out method with these classifiers.

5.1 FISHER CLASSIFIER

Suppose that there are two classes. The training patterns from classes 1 and 2 are

$$X_1^1, X_2^1, \cdots, X_{N_1}^1 ; \; X_1^2, X_2^2, \cdots, X_{N_2}^2$$

The means and covariance matrices of the patterns in the classes are estimated as

$$\hat{m}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_j^i,\ \hat{\Sigma}_i = \frac{1}{(N_i - 1)}\sum_{j=1}^{N_i}(X_j^i - \hat{m}_i)(X_j^i - \hat{m}_i)^T,\quad i = 1, 2$$

The Fisher weight vectors are given by

$$W_i = \begin{bmatrix} \hat{S}_W^{-1}\hat{m}_i \\ -\hat{m}_i^T\hat{S}_W^{-1}\ \dfrac{(\hat{m}_1 + \hat{m}_2)}{2} \end{bmatrix} = \begin{bmatrix} V_i \\ v_i \end{bmatrix},\ i = 1, 2$$

where $\hat{S}_W = \hat{\Sigma}_1 + \hat{\Sigma}_2$. The Fisher's decision rule is as follows:

$$\text{decide } X \in \omega_1 \text{ if } V_1^T X + v_1 > V_2^T X + v_2$$

$$\text{decide } X \in \omega_2 \text{ otherwise}$$

For use with the leave-one-out method, recursive expressions for computing the weight vectors are given below. Let a pattern X_k^1 from class ω_1 be left out. Define the means and covariance matrices of the total pattern set as follows:

$$\hat{m}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_j^i,\ i = 1, 2$$

$$\hat{\Sigma}_1 = \frac{1}{(N_1 - 2)}\sum_{j=1}^{N_1} (X_j^1 - \hat{m}_1)(X_j^1 - \hat{m}_1)^T$$

$$\hat{\Sigma}_2 = \frac{1}{(N_2 - 1)}\sum_{j=1}^{N_2} (X_j^2 - \hat{m}_2)(X_j^2 - \hat{m}_2)^T$$

Let $\hat{S}_W = \hat{\Sigma}_1 + \hat{\Sigma}_2$. Compute V_i, v_i, $i = 1, 2$ as

$$V_1 = \hat{S}_W^{-1}\hat{m}_1 \qquad\qquad V_2 = \hat{S}_W^{-1}\hat{m}_2$$

$$v_1 = \frac{-\hat{m}_1^T\hat{S}_W^{-1}(\hat{m}_1 + \hat{m}_2)}{2} \qquad\qquad v_2 = \frac{-\hat{m}_2^T\hat{S}_W^{-1}(\hat{m}_1 + \hat{m}_2)}{2}$$

Note that $\hat{\Sigma}_1$ is defined differently from the usual unbiased estimate for covariance matrices. It is defined thus for mathematical simplicity and will not affect the results. The weight vectors when a pattern X_k^1 from class ω_1 was left out become

$$V_1(X_k^1) = \hat{S}_{W1k}^{-1}\hat{m}_{1k}$$

$$v_1(X_k^1) = \frac{-\hat{m}_{1k}^T\hat{S}_{W1k}^{-1}(\hat{m}_{1k} + \hat{m}_2)}{2}$$

$$V_2(X_k^1) = \hat{S}_{W1k}^{-1}\hat{m}_2$$

$$v_2(X_k^1) = \frac{-\hat{m}_2^T\hat{S}_{W1k}^{-1}(\hat{m}_{1k} + \hat{m}_2)}{2}$$

where

$$\hat{m}_{1k} = \frac{1}{(N_1 - 1)}\sum_{\substack{j=1\\ \neq k}}^{N_1} X_j^1$$

$$\hat{\Sigma}_{1k} = \frac{1}{(N_1 - 2)}\sum_{\substack{j=1\\ \neq k}}^{N_1} (X_j^1 - \hat{m}_{1k})(X_j^1 - \hat{m}_{1k})^T$$

$$\hat{S}_{W1k} = \hat{\Sigma}_{1k} + \hat{\Sigma}_2$$

The $\hat{m}_2$ and $\hat{\Sigma}_2$ are defined as before. With these, $V_i(X_k^1)$ and $v_i(X_k^1)$ can be expressed in terms of V_i and v_i as follows.

Let

$$\alpha = \frac{N_1}{(N_1 - 1)(N_1 - 2)}$$

$$\beta(X_k^1) = (X_k^1 - \hat{m}_1)^T \hat{S}_W^{-1} (X_k^1 - \hat{m}_1)$$

$$\nu(X_k^1) = 1 - \alpha\beta(X_k^1)$$

$$Y_1 = \hat{S}_W^{-1} (X_k^1 - \hat{m}_1)$$

$$d_1 = (X_k^1 - \hat{m}_1)^T \hat{S}_W^{-1} \hat{m}_1 = (X_k^1 - \hat{m}_1)^T V_1$$

$$\hat{m} = \frac{(\hat{m}_1 + \hat{m}_2)}{2}, \; d_2 = Y_1^T \hat{m}$$

Then.

$$V_1(X_k^1) = V_1 + \alpha \frac{d_1}{\nu(X_k^1)} Y_1 - \frac{1}{(N_1 - 1)\nu(X_k^1)} Y_1$$

$$v_1(X_k^1) = v_1 - \frac{\alpha}{\nu(X_k^1)} d_1 d_2 + \frac{1}{(N_1 - 1)} d_2 + \frac{\alpha\beta(X_k^1)}{(N_1 - 1)\nu(X_k^1)} d_2 + \frac{1}{2(N_1 - 1)} d_1$$
$$+ \frac{\alpha\beta(X_k^1)}{2(N_1 - 1)\nu(X_k^1)} d_1 - \frac{1}{2(N_1 - 1)^2} \frac{\beta(X_k^1)}{\nu(X_k^1)}$$

Let

$$d_3 = (X_k^1 - \hat{m}_1)^T V_2$$

Then

$$V_2(X_k^1) = V_2 + \frac{\alpha Y_1 d_3}{\nu(X_k^1)}$$

$$v_2(X_k^1) = v_2 + \frac{1}{2(N_1 - 1)} d_3 - \frac{\alpha d_3 d_1}{\nu(X_k^1)} + \frac{1}{2(N_1 - 1)} \frac{\alpha d_3 \beta(X_k^1)}{\nu(X_k^1)}$$

5.2 ERROR CORRECTION CLASSIFIER

Another classifier selected for evaluation is a candidate from a nonparametric family of classifiers. It is an error correction classifier that uses linear discriminant functions for each class.

$$g_i(X) = W_i^T X, \; i = 1, 2, \cdots, M$$

The decision rule is to decide a pattern $X \in \omega_i$, if

$$g_i(X) > g_j(X), \; j \underset{\neq i}{=} 1, 2, \cdots, M$$

A brief description of the algorithm is as follows. Suppose that training patterns $X_1^i, X_2^i, \cdots, X_{N_i}^i$; $i = 1, 2, \cdots, M$ are given. From the training patterns of class ω_i, form a matrix $A(i)$,

$$A(i) = \begin{bmatrix} X_1^{i^T} \\ X_2^{i^T} \\ \vdots \\ X_{N_i}^{i^T} \end{bmatrix}, \; i = 1, 2, \cdots, M$$

Let

$$A = \begin{bmatrix} A(1) \\ A(2) \\ \vdots \\ A(M) \end{bmatrix}, \; W = [W_1, W_2, \cdots, W_M]$$

For a particular initial matrix B(0), form

$$AW(0) = B(0)$$

and obtain initial weight vector matrix W(0),

$$W(0) = (A^TA)^{-1}A^TB(0)$$

Adjust B()'s and W()'s simultaneously until the correction process stops or for a predetermined number of iterations and take the weight vector matrix at that point as the solution weight vector matrix.

The implementation of the leave-one-out method with this classifier is as follows. Compute the B-matrix with the solution weight vector matrix at the point of stopping. That is,

$$B = AW$$

Let a pattern X from class ω_M be left out. Let the corresponding row of B-matrix be C. That is,

$$A = \begin{bmatrix} A_{N-1} \\ \hdashline X \end{bmatrix}; \quad B = \begin{bmatrix} B_{N-1} \\ \hdashline C \end{bmatrix}$$

where $C^T = [c^1, c^2, \cdots, c^M]$. Let

$$W = (A^TA)^{-1}A^TB$$

and

$$W_{N-1} = (A_{N-1}^TA_{N-1})^{-1}A_{N-1}^TB_{N-1}$$

The relationship between the matrices W and W_{N-1} can be derived as

$$W_{N-1} = W + \frac{(A^TA)^{-1}X(X^TW - C^T)}{(1 - \theta)}$$

where $\theta = X^T(A^TA)^{-1}X$. Every time a pattern is left, a new weight vector matrix is recursively computed and the left-out pattern is tested, and the biases can be computed.

5.3 CLASSIFIER WITH LABELING UNCERTAINTIES

The training patterns and classes are described in section 5.2. Let a_j^i be an M-dimensional vector associated with the pattern X_j^i; its ith component could be the probability that the label of the pattern X_j^i is the ith class.

In the absence of this knowledge, if the training pattern comes from class i, the ith component of a_j^i could be set to 1 and the rest of the components to zero. The a_j^i can also be set so as to map the patterns of each class into vertices of a simplex.

It is proposed to use the weight vectors that minimize the mapping errors in the minimum mean square error sense to the vectors a_j^i. Then the criterion is

$$C = \sum_{i=1}^{M} \sum_{j=1}^{N_i} (W^T X_j^i - a_j^i)^T (W^T X_j^i - a_j^i)$$

The weight vector matrix W that minimizes this criterion is given by

$$W = S^{-1} B$$

where

$$S = \sum_{i=1}^{M} \sum_{j=1}^{N_i} X_j^i X_j^{i^T}$$

and

$$B = \sum_{i=1}^{M} \sum_{j=1}^{N_i} X_j^i a_j^{i^T}$$

Suppose a pattern X_k^i from class ω_i is left. Then the weight vector matrix $W(X_k^i)$ is related to W as

$$W(X_k^i) = W + \frac{Y_k^i X_k^{i^T} W}{1 - X_k^{i^T} Y_k^i} - \frac{1}{1 - X_k^{i^T} Y_k^i} Y_k^i a_k^{i^T}$$

where

$$Y_k^i = S^{-1} X_k^i$$

6. INCORPORATION OF SPATIAL INFORMATION

It is proposed to incorporate the spatial information into the classification through transition probabilities. The dependencies in adjacent pixels are modeled as follows. If I and J are neighboring pixels,

$$P(I = \omega_i | J = \omega_i) = (1 - \theta)P(I = \omega_i) + \theta$$

and

$$P(I = \omega_i | J = \omega_j) = (1 - \theta)P(I = \omega_i)$$

where the parameter θ controls the dependencies between neighboring pixels. The $\theta = 1$ represents the complete dependence, and $\theta = 0$ represents the complete independence. The following two models that consider spatial information will be investigated. The posteriori probabilities for use with these algorithms will be estimated from the outputs of linear classifiers.

6.1 TWO-DIMENSIONAL SPATIALLY UNIFORM CONTEXT

Consider a neighborhood of 9 pixels shown in the following figure.

8	1	2
7	0	3
6	5	4

Figure 1.– Illustration of 3×3 neighborhood.

Suppose that the pixel 0 is under consideration and that pixels 1 through 8 are its neighbors. Decide the pixel $X_0 \in \omega = \ell$ that maximizes

$$P(X_0|\omega = \ell)P(\omega = \ell) \prod_{i=1}^{8} [\theta p(X_i|\omega = \ell) + (1 - \theta) \sum_{j=1}^{M} p(X_i|\omega = j)P(\omega = j)]$$

6.2 SEQUENTIAL OR MARKOVIAN DEPENDENCE

This section considers the sequential Markovian dependence between neighboring pixels with the transition probabilities described in section 6.1 in terms of parameter θ. Sequential look-before and look-ahead type of context will be used with the classifier:

X_1, ω_1	X_2, ω_2	X_3, ω_3	...	X_{n-1}, ω_{n-1}	X_n, ω_n	X_{n+1}, ω_{n+1}	...	

Figure 2.— Illustration of pixels and labels on scan line.

Suppose we want to find the label ω_n of pattern X_n, using look-before and look-ahead type of context. The posteriori probabilities of ω_n using the context are given by

$$p(\omega_n = k|X_1, \cdots, X_{n-1}, X_n) = \frac{P(X_n|\omega_n = k)\left[(1 - \theta)P(\omega_n = k) + \theta p(\omega_{n-1} = k|X_1, \cdots, X_{n-1})\right]}{\left[\sum_{j=1}^{M} p(X_n|\omega_n = j)\left\{(1 - \theta)P(\omega_n = j) + \theta p(\omega_{n-1} = j|X_1, \cdots, X_{n-1})\right\}\right]}$$

$$p(\omega_n = k|X_1, \cdots, X_n, X_{n+1}) = \frac{p(\omega_n = k|X_1, \cdots, X_n)\left[(1 - \theta)\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j) + \theta p(X_{n+1}|\omega_{n+1} = k)\right]}{\sum_{i=1}^{M} p(\omega_n = i|X_1, \cdots, X_n)\left[(1 - \theta)\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j) + \theta p(X_{n+1}|\omega_{n+1} = k)\right]}$$

This sequential algorithm will be applied to find the label of a center pixel 0 in a 3×3 neighborhood as follows.

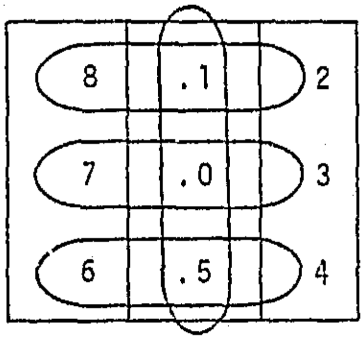

Figure 3.— Sequential contextual algorithm for a 3×3 neighborhood.

Pixels 8, 1, and 2 will be used to find the posteriori probabilities for pixel 1 and similarly for pixels 7, 0, 3 and for pixels 6, 5, 4. Finally, pixels 1, 0, 5 will be used to find the label of the pixel under consideration, 0.

7. CLASSIFIER DESIGN WITH IMPERFECT LABELS

The techniques for handling imperfections in the labels developed in references 8 and 9 will be implemented and evaluated. To conduct the investigations proposed in this plan, the required computer time will be at least 45 computer processing unit hours. This time allows for generating the necessary data files. It is estimated that this task requires two people over a period of 6 months. Software support is required for developing necessary software.

8. REFERENCES

1. Chittineni, C. B.: On the Dimensionality and Sample Size for Feature Selection in Pattern Recognition. Tech. rpt., 3M Company (St. Paul, Minn.), 1975.

2. Foley, D. H.: Considerations of Sample and Feature Size. IEEE Trans. Information Theory, Sept. 1972, pp. 618-626.

3. Lennington, R. K.: Research Plan for Clustering Algorithm Evaluation. LEC-13296, to be published.

4. Chittineni, C. B.: On the Fisher Criterion and Divergence in Pattern Recognition. Internat. J. Control, vol. 16, no. 1, July 1972, pp. 205-207.

5. Chittineni, C. B.: Estimation of Probability of Error. Pattern Recognition, vol. 9, no. 4, 1977, pp. 191-196.

6. Chittineni, C. B.: Pattern Classification Algorithms. Tech. rpt., 3M Company (St. Paul, Minn.), 1975.

7. Ho, Y. C.; and Kashyap, R. L.: An Algorithm for Linear Inequalities and Its Applications. IEEE Trans. Electronic Computers, Oct. 1965, pp. 683-688.

8. Chittineni, C. B.: Learning with Imperfectly Labeled Patterns. Paper submitted for presentation at IEEE Computer Society Conference on Pattern Recognition and Image Processing (Chicago, Illinois), Aug 6-8, 1979.

9. Chittineni, C. B.: On the Selection of Effective Features from the Imperfectly Labeled Patterns. Internat. J. Computer and Information Sciences, vol. 2, no. 2, 1973, pp. 103-114.